Specification

Voice Processing Unit and System, and

Voice Processing Method

5   Technical Field

The present invention relates to a voice

processing technique and, more particularly, to a

system, unit, and method which transmit voice

information input on a terminal (client) side to a voice

10 processing unit through a network and process the

information.

Background Art

As a conventional system of this type, there

is known a technique which makes a cell phone terminal

15 (client) phone-connect to a voice processing server by

using a phone-to function or the like, performs voice

processing (voice recognition, speaker collation, and

the like) for voice uttered by a user, transmits the

result from the voice processing server to a Web server,

20 makes the Web server generate a window reflecting the

processing result, and makes the cell phone terminal

download and display the window, thereby associating

voice processing with window within this framework (see,

for example, Japanese Patent No. 3452250 (reference 1)).

25 As shown in Fig. 1, in this conventional system, a cell

phone terminal 11 and a voice processing server 13

transmit/receive data through a circuit switched network

- 1 -

15, and the cell phone terminal 11 and a Web server 12 transmit/receive data through a packet network 14.

If the Web server 12 and the voice processing server 13 have received access from a plurality of cell phone terminals 11, a technique of comprehending the relationship between a window downloaded from the Web server 12 to the cell phone terminal 11 and voice data transmitted from the cell phone terminal 11 to the voice processing server 13 is necessary to make the cell phone terminal 11 reflect a voice processing result in the window and display it.

The conventional system shown in Fig. 1 is configured to allow the Web server 12 and the voice processing server 13 to uniquely comprehend a terminal which downloads window information and a terminal which transmits voice data by associating the terminal ID of the cell phone terminal 11 with a cell phone terminal number.

There is also known a technique of transmitting feature vectors and voice information such as compressed voice data from a client such as a portable digital assistant (PDA) or on-vehicle terminal to a voice processing server through a packet network so as to perform voice processing (voice recognition, speaker collation, and the like) (see, for example, Japanese Patent Laid-Open No. 2003-5949 (reference 2)).

The system disclosed in reference 2 can

operate contents designed to display a processing result in the form of a table and display the result obtained by a search based on a processing result in a window.

Disclosure of Invention

Problem to be Solved by the Invention

The system disclosed in reference 2 requires a technique of allowing a server side to comprehend the relationship between a window downloaded to a client and voice data transmitted from the client even within a framework for voice processing which is designed to transmit/receive data through a packet network.

The conventional technique disclosed in reference 1 is a method of associating a phone number with the terminal ID of a cell phone terminal, and hence is a technique which cannot be used in the above framework for voice processing which uses a packet network which does not require any phone numbers. This makes it necessary to use a new technique of allowing the server side to comprehend the relationship between a window downloaded to a client and voice data transmitted from the client in a framework for voice processing in which data is transmitted/received between the client, a voice processing server, and a Web server through a packet network.

It is, therefore, an object of the present invention to allow a server side to comprehend the relationship between information downloaded from an

information providing server (information providing unit) such as a Web server to a client (terminal) and voice information transmitted from the client to a voice processing server (voice processing unit).

5   It is another object of the present invention to download proper information reflecting a voice processing result even if a voice processing server and an information providing server receive access from a plurality of clients.

10   Means of Solution to the Problem

In order to achieve this object, according to the present invention, there is provided a voice processing system characterized by comprising a terminal which transmits input voice information and outputs

15   received information, a voice processing unit which performs voice processing on the basis of voice information from the terminal, and an information providing unit which receives a voice processing result obtained by the voice processing unit and transmits

20   information reflecting the voice processing result to the terminal, wherein the terminal, the voice processing unit, and the information providing unit share processing identification information corresponding to a series of processes performed by the voice processing

25   unit and the information providing unit on the basis of the voice information.

In addition, according to the present

invention, there is provided a voice processing method characterized by comprising the steps of causing a terminal to transmit input voice information to a voice processing unit, causing the voice processing unit to

5   perform voice processing for the voice information from the terminal, transmitting a voice processing result to an information providing unit, and causing the information providing unit to prepare information reflecting the voice processing result obtained by the

10 voice processing unit, and the step of transmitting the prepared information to the terminal, wherein the terminal, the voice processing unit, and the information providing unit share processing identification information corresponding to a series of processes

15 performed by the voice processing unit and the information providing unit on the basis of the voice information.

According to the present invention, there is provided an information providing server unit

20 characterized by comprising first reception means for receiving a service request signal from a client, identification information generating means for generating processing identification information corresponding to a series of processes performed on the

25 basis of voice information from the client when the service request signal is received, means for generating first information to be presented to the client on the

basis of the processing identification information, first transmission means for transmitting the processing identification information and the first information to the client, second reception means for receiving a voice

5 processing result and the processing identification information from a voice processing server which performs voice processing upon receiving the voice signal and the processing identification information from the client, means for generating second information

10 reflecting the voice processing result in correspondence with the processing identification information from the voice processing server, and second transmission means for transmitting the second information to the client.

According to the present invention, there is

15 provided a client unit characterized by comprising unique identification information output means for outputting unique identification information of the client unit as processing identification information corresponding to a series of processes performed by a

20 voice processing server which performs voice processing for voice information from the client unit and an information providing server which transmits information reflecting a voice processing result obtained by the voice processing server to the client unit, first

25 transmission means for transmitting a service request signal and the processing identification information to the information providing server when a service request

is issued, and second transmission means for transmitting the input voice information to the voice processing server together with the processing identification information.

5   According to the present invention, there is provided a voice processing server unit characterized by comprising first reception means for receiving a voice processing request signal from a client, identification information generating means for generating processing

10   identification information corresponding to a series of processes performed on the basis of voice information from the client when the voice processing request signal is received, first transmission means for transmitting the processing identification information to the client,

15   second reception means for receiving the voice information and the processing identification information from the client, voice processing executing means for performing voice processing for the voice information from the client, and transmission means for

20   transmitting, to an information providing server, a voice processing result obtained by the voice processing executing means and the processing identification information from the client, while generating information reflecting the voice processing result in

25   correspondence with the processing identification information.

          A program according to the present invention

is a program for causing a computer forming the above information providing server unit, the above client unit, or the above voice processing server unit to implement the functions of the respective units.

5   An information processing system according to the present invention is characterized by comprising a client and a plurality of servers,

wherein a series of processes (A), (B), and (C):

10   (A) in association with processing executed by at least one of the plurality of servers on the basis of a request from the client, processing is performed by another server in accordance with the request,

(B) exchanging a processing result between

15   another server and one server, and

(C) causing one server to generate response information in response to the request on the basis of the processing result

are managed by common processing identification

20   information shared by the client, one server, and another server.

Effects of the Invention

According to the present invention, a client (terminal), a voice processing server (voice processing

25   unit), and an information providing server (information providing unit) share processing identification information corresponding to a series of processes

performed by the voice processing server and the information providing server on the basis of voice information, thereby allowing the server side to comprehend the relationship between information

5   downloaded from the information providing server to the client and voice information transmitted from the client to the voice processing server. As a consequence, even if the voice processing server and the information providing server receive access from a plurality of

10 clients, the user can download proper information reflecting a voice processing result.

        This makes it possible to provide a content which associates voice processing with a window, e.g., displaying, on a window, the result obtained by

15 performing processing such as a search performed on the basis of voice information uttered by the user or downloading proper information on the basis of voice information uttered by the user.
Brief Description of Drawings

20   Fig. 1 is a view showing the arrangement of a conventional system;

        Fig. 2 is a view showing the arrangement of an embodiment of the present invention;

        Fig. 3 is a view showing the arrangement of

25 the first embodiment of the present invention;

        Fig. 4 is a view showing the arrangement of the second embodiment of the present invention;

Fig. 5 is a view showing the arrangement of the third embodiment of the present invention;

Fig. 6 is a view showing the arrangement of a client according to the first specific example of the present invention;

Fig. 7 is a view showing the arrangement of a Web server according to the first specific example of the present invention;

Fig. 8 is a view showing the arrangement of a voice processing server according to the first specific example of the present invention;

Fig. 9 is a view showing the arrangement of a client according to the second specific example of the present invention;

Fig. 10 is a view showing the arrangement of a Web server according to the second specific example of the present invention;

Fig. 11 is a view showing the arrangement of a voice processing server according to the third specific example of the present invention;

Fig. 12 is a view for explaining the operation of the first specific example of the present invention;

Fig. 13 is a view for explaining the operation of the second specific example of the present invention;

Fig. 14 is a view for explaining the operation of the third specific example of the present invention;

Fig. 15 is a view for explaining an example of

transition to a window (page) displayed on the client according to the first specific example of the present invention; and

Fig. 16 is a view for explaining another
5 example of transition to a window (page) displayed on the client according to the first specific example of the present invention.

Best Mode for Carrying Out the Invention

The embodiments of the present invention will
10 be described in detail below with reference to the accompanying drawings.

Referring to Fig. 2, according to an embodiment of the present invention, a client (terminal) 10, a Web server (an information providing server or
15 information providing unit) 20, and a voice processing server (voice processing unit) 30 are connected through a network. The client 10 comprises a voice data input unit, a browser function, and a communication function of connecting to a packet network 40 such as an IP
20 network as a network. The client 10, Web server 20, and voice processing server 30 share processing identification information corresponding to a series of processes performed by the Web server 20 and the voice processing server 30 on the basis of voice data. As
25 processing identification information, for example, an ID (to be referred to as a "session ID") assigned in correspondence with a session of utterance processing or

a unique ID held by the client 10 is used. Sharing such
processing identification information makes it possible
to comprehend the correspondence relationship between a
window downloaded from the Web server 20 to the client
5   10 and voice data transmitted from the client 10 to the
voice processing server 30.
[First Embodiment]
        Fig. 3 is a view showing the arrangement of
the first embodiment of the present invention, in which
10   a Web server 20 comprises a session ID generating unit
which generates a session ID for each session.
        A procedure for processing in this embodiment
will be described with reference to Fig. 3. When a
client 10 issues a request for a service using voice
15   processing to the Web server 20, the Web server 20
generates a session ID.
        When the client 10 downloads window
information from the Web server 20, the Web server 20
transmits the generated session ID to the client 10.
20   For example, the session ID may be transmitted while,
for example, being contained in window information.
        When transmitting the voice information of
input voice to a voice processing server 30, the client
10 transmits the session ID received from the Web server
25   20 to the voice processing server 30. The ID may be
transmitted while being contained in the voice
information or may be transmitted separately.

The voice processing server 30 performs voice

processing (voice recognition, speaker collation, and

the like) on the basis of the received voice

information. The voice processing server 30 transmits

5   the session ID when transmitting the voice processing

result to the Web server 20. A session ID may be

transmitted while being contained in a voice processing

result.

The Web server 20 can associate the voice

10 processing result obtained by the voice processing

server 30 with the client 10 which has issued a service

request in accordance with session ID information, and

allows the client 10 to download a window reflecting the

processing result. In this case, the Web server 20 may

15 be configured to transmit a window (page) containing

voice processing resultant information such as the voice

recognition result of an utterance or the like to the

client 10 and to download window information

corresponding to the voice processing result upon

20 selection from the client 10.

[Second Embodiment]

Fig. 4 is a view showing the arrangement of

the second embodiment of the present invention, which

comprises an arrangement which uses the ID held by a

25 client 10 as a unique ID. A processing procedure will

be described, in a case wherein the ID held in advance

by the client 10 is to be used as an ID (unique ID)

- 13 -

unique to the client or an ID (unique ID) unique to the client is to be generated by using the ID held in advance by the client 10.

When issuing a request for a service using voice processing to a Web server 20, the client 10 notifies the Web server 20 of the ID held in advance by itself as a unique ID. Alternatively, the client 10 newly generates an ID unique to the client by using the ID held in advance by itself, and notifies the Web server 20 of the generated unique ID. For example, a unique ID may be generated by assigning time stamp information to the ID held in advance by itself.

Subsequently, window information of the requested service is downloaded from the Web server 20 to the client 10.

The window downloaded from the Web server 20 is displayed on a window display unit 140 of the client 10. The client 10 receives the voice signal input from the user and converts it into voice information. When transmitting the voice information to a voice processing server 30, the client 10 also transmits the unique ID.

The voice processing server 30 performs voice processing on the basis of the received voice information. When transmitting the voice processing result to the Web server 20, the voice processing server 30 also transmits the unique ID to the Web server 20.

The Web server 20 receives the voice

processing result and the unique ID from the voice

processing server 30. The Web server 20 can associate

the voice processing result with the client 10, from

which the service request has been received, in

5   accordance with the unique ID from the voice processing

server 30, and allows the client 10 to download window

information reflecting the voice processing result. In

this case, the Web server 20 is configured to transmit a

window (page) containing the voice processing resultant

10   information such as the voice recognition result on

voice and the like to the client 10, and to download

window information corresponding to the voice processing

result in accordance with selection by the client 10.

[Third Embodiment]

15   Fig. 5 is a view showing the arrangement of

the third embodiment of the present invention, in which

a voice processing server 30 comprises a session ID

generating unit which generates a session ID generated

for each session. A processing procedure according to

20   this embodiment will be described with reference to

Fig. 5. When a client 10 accesses the voice processing

server 30 for the transmission of voice information, a

session ID generating unit 31 of the voice processing

server 30 generates a session ID and notifies the client

25   10 of it.

The client 10 then notifies a Web server 20 of

the received session ID.

The voice processing server 30 performs voice processing on the basis of the voice information received from the client 10. When transmitting the voice processing result to the Web server 20, the voice

5 processing server 30 also transmits the session ID to the Web server 20.

The Web server 20 can associate the voice processing result with the client from which the service request has been received in accordance with the session

10 ID information, and allows the client 10 to download a window reflecting the processing result. In this case, the Web server 20 may be configured to transmit a window (page) containing the voice processing resultant information such as the voice recognition result on

15 voice or the like to the client 10 and to download window information corresponding to the voice processing result upon selection from the client 10.

In the embodiment shown in Fig. 3, the Web server 20 may transmit a session ID to the client 10 in

20 the following manner:

· embedding the session ID as tag information in a window (HTML, XML, or the like) or

· embedding the session ID as header information.

In each embodiment described with reference to

25 Figs. 3 to 5, the client 10 may transmit a session ID to the voice processing server 30 by the following method:

· embedding the session ID as header information in a

packet of voice information or

· embedding the session ID as part of voice

information.

In each embodiment described above with

5   reference to Figs. 3 to 5, the voice processing server

30 may transmit a session ID to the Web server 20 by the

following method:

· transmitting the session ID as header information of

a packet of voice processing resultant information or

10   · containing the session ID as part of a voice

processing result.

The present invention will be described more

in detail by way of specific examples.

[First Specific Example]

15   The first specific example of the voice

processing system according to the present invention

will be described with reference to Fig. 2. The client

10 is connected to the Web server 20 and the voice

processing server 30 through the network (packet

20   network) 40. The following can be enumerated as

clients: a portable terminal, PDA (Personal Digital

Assistant), on-vehicle terminal, PC (Personal Computer),

home terminal, and the like. For example, the following

can be enumerated as the Web server 20 and the voice

25   processing server 30: a computer installed with Windows

XP (registered trademark), Windows 2000 (registered

trademark), or the like as an OS (Operating System) and

a computer installed with Solaris (registered trademark) as an OS. As the network (packet network) 40, an IP network such as the Internet (wired/wireless) or an intranet is used.

5   In this specific example, the Web server 20 includes a session ID generating unit which generates a session ID.

Fig. 6 is a view showing the arrangement of the client 10 according to the first specific example of the present invention. Referring to Fig. 6, the client

10   10 comprises a data input unit 110 which functions as a voice input unit and inputs voice data, a window display unit 140, a data communication unit 130, and a control unit 120.

15   Fig. 7 is a view showing the arrangement of the Web server 20. Referring to Fig. 7, the Web server 20 comprises a data communication unit 210, content management unit (information management means) 220, and session ID generating unit 230.

20   Fig. 8 is a view showing the arrangement of the voice processing server 30. Referring to Fig. 8, the voice processing server 30 comprises a data communication unit 310, control unit 320, and voice processing executing unit 330.

25   Fig. 12 is a view for explaining the sequence operation of this specific example. The specific example will be described with reference to Figs. 6 to 8

and 12.

The client 10 issues a request for a service including voice processing to the Web server 20 (step S101). More specifically, when a button on the window displayed on the client 10 is clicked, a service request signal is transmitted to the Web server 20. The Web server 20 then activates a program such as a CGI (Common Gateway Interface) which executes the service.

In the Web server 20, the data communication unit 210 receives the service request signal from the client 10 (step S201), and transmits it to the content management unit 220.

The content management unit 220 checks the service and then transmits the service request signal to the session ID generating unit 230. The session ID generating unit 230 receives the service request signal and generates a session ID (step S202). For example, a session ID may be generated by counting up a predetermined initial value by a value corresponding to an access count.

The generated session ID is transmitted to the content management unit 220. The content management unit 220 generates a window to be downloaded to the client 10 on the basis of the received session ID (step S203). This window may be generated by containing the session ID in URL (Uniform Resource Locator) information linked to a button for result acquisition.

The content management unit 220 of the Web server 20 then downloads the generated window to the client through the data communication unit 210 of the Web server 20 (step S204). At this point of time, the

5   Web server 20 also transmits the session ID to the client 10. The session ID may be transmitted by the following method:

· writing the session ID as tag information in the window generated by the Web server 20 or

10 · writing the session ID as a header of a packet.

In the client 10, the data communication unit 130 receives the window information and the session ID from the Web server 20 (step S102) and transmits them to the control unit 120 of the client 10. The window

15 information is transmitted from the control unit 120 to the window display unit 140 to be displayed. The window information displayed on the client 10 includes, for example, selection/prompt for voice input or the like by the user.

20   The voice uttered by the user is input to the data input unit 110 of the client 10 (step S104) and is transmitted to the control unit 120 in the client 10. The control unit 120 of the client 10 performs necessary data processing (step S105). Data processing to be

25 performed includes, for example, digitalization processing for input voice, voice detection processing, voice analysis processing, and voice compression

processing. As voice data, for example, digitalized voice data, compressed voice data, or a feature vector is used (for details see Seiichi Nakagawa, "Voice Recognition by Probability Model", THE INSTITUTE OF ELECTRONICS, INFORMATION AND COMMUNICATION ENGINEERS (reference 3), pp. 10 - 12).

In this data processing, a session ID is contained in voice data. More specifically, this processing may be performed by

· containing the session ID as header information of a voice data packet or

· containing the session ID as part of voice data.

The data communication unit 130 sequentially transmits the data processed by the control unit 120 of the client 10 to the voice processing server 30.

In the voice processing server 30, the data communication unit 310 receives the data sequentially transmitted from the client (step S301). If the control unit 320 determines that the received data is voice data, the control unit 320 transmits the data to the voice processing executing unit 330.

The voice processing executing unit 330 comprises at least one of the following (not shown) required for voice processing: a recognition engine, recognition dictionary, synthesis engine, synthesis dictionary, and speaker collation engine, and sequentially performs voice processing (step S302).

Note that the contents of voice processing change depending on the type of data to be transmitted from the client 10. If, for example, the data to be transmitted is compressed voice data, the voice processing executing unit 330 performs decompression of the compressed data, voice analysis, and matching processing. If a feature vector is to be transmitted from the client 10, only matching processing is performed.

When the voice processing is complete, the voice processing executing unit 330 of the voice processing server 30 transmits the voice processing result to the data communication unit 310 through the control unit 320. The data communication unit 310 then transmits the result to the Web server 20 (step S303).

The voice processing result transmitted from the voice processing server 30 to the Web server 20 contains at least one of recognition resultant information, speaker collation information, voice (synthetic voice, voice obtained by converting the input voice, or the like), and the like. At this time, the voice processing server 30 also transmits the session ID to the Web server 20. The session ID may be transmitted by the following method:

· containing the session ID as header information of a packet for the transmission of the voice processing result or

·   transmitting the session ID as part of the voice

processing result.

In the Web server 20, the data communication

unit 210 receives the voice processing result and the

5   session ID (step S205) and transmits them to the content

management unit 220.

The content management unit 220 generates, for

each session ID, resultant information based on the

voice processing result (e.g., voice recognition

10   resultant information; see a window 1003 in Figs. 15 and

16 (to be described later) or the like) or content

information reflecting the voice processing result (a

window, voice, moving image, or the like) (step S206).

The Web server 20 downloads the resultant

15   information and a content or only the content, generated

for each session ID, to the client 10 which has issued

the service request (step S207).   The client 10 then

receives the downloaded resultant information/content

(step S106).

20   More specifically, at the start of voice

processing, a URL linked to a result acquisition button

on the window downloaded from the Web server 20 to the

client 10 is formed into a URL containing the session

ID.   The content management unit 220 then places the

25   content information in which the voice processing result

is reflected by the Web server 20 at the place

represented by the URL containing the session ID.   When

the user presses the result acquisition button (e.g., the "display map" button on the window 1003 in Fig. 15) of the client 10, the URL containing the session ID is designated, and the content information (e.g., the map window on a window 1004 in Fig. 15) corresponding to this URL is downloaded.

More specifically, this technique can be used for various kinds of processing in, for example, the following cases:

· a case wherein the client 10 interacts with the voice processing server 30,

· a case wherein the voice processing server 30 performs a search or the like by using the voice processing result, or

· a case wherein the Web server 20 performs a search or the like by using the voice processing result.

Note that the processing to be performed by the client 10, Web server 20, and voice processing server 30, which is exemplified by Fig. 12, i.e., the functions of them, may be implemented by programs executed on computers forming the client 10, Web server 20, and voice processing server 30. In addition, the present invention may implement the Web server 20 and the voice processing server 30 on one computer or on remote computers. In this case, when the Web server 20 and the voice processing server 30 exchange IDs, the arguments of a subroutine call may be used.

Alternatively, when the Web server 20 and the voice processing server 30 are to exchange variables by inter-process communication, the variables to be exchanged may be those referred to commonly. In addition, in this specific example, the present invention can be applied to a system in which a client which issues a processing request to a server is implemented on the same computer as that on which the server is implemented. That is, the present invention can be applied to an arbitrary management system for executing a request from a client by making a plurality of servers operate in cooperation with each other.

[Second Specific Example]

The second specific example of the present invention will be described next. More specifically, as shown in Fig. 4, when the ID held by the client 10 is to be used as an ID (unique ID) unique to the client, an ID (unique ID) unique to the client is generated by using the ID held in advance by the client.

Fig. 9 is a view showing the arrangement of the client 10 according to the second specific example of the present invention. Referring to Fig. 9, the client 10 comprises a data input unit 110 which functions as a voice input unit and inputs voice data, a window display unit 140, a data communication unit 130, a control unit 120, and a unique ID holding/generating unit (unique identification information output means)

150.

Fig. 10 is a view showing the arrangement of the Web server 20. Referring to Fig. 10, the Web server 20 comprises a data communication unit 210 and a content management unit 220.

The voice processing server 30 has the arrangement shown in Fig. 8, and comprises a data communication unit 310, control unit 320, and voice processing executing unit 330.

Fig. 13 is a view for explaining the sequence operation of this specific example. The specific example will be described with reference to Figs. 9, 10, 8, and 13.

When issuing a request for a service using voice processing to the Web server 20, the client 10 transmits, as a unique ID (unique identification information), the ID (terminal identification information) held in advance by the client 10 to the control unit 120 (step S111). Alternatively, in step S111, the unique ID holding/generating unit 150 generates an ID unique to the client by using the ID held in advance and notifies the control unit 120 of the generated unique ID. A unique ID may be generated by assigning time stamp information to the ID held in advance. The control unit 120 receives the service request and the ID and transmits the received unique ID to the Web server 20 through the data communication unit

130 (step S112).

In the Web server 20, the data communication unit 210 receives the service request signal containing the received voice processing and the unique ID (step S211). The data communication unit 210 transmits the service request signal and the unique ID to the content management unit 220.

After checking the service, the content management unit 220 generates a window (first information) to be downloaded to the client 10 by using the received unique ID (step S212). The window may be generated in the following manner (by the following method). As in the above specific example, a session ID is contained in URL (Uniform Resource Locator) information linked to a result acquisition button.

The window generated by the content management unit 220 is downloaded to the client 10 through the data communication unit 210 (step S213).

In the client 10, the data communication unit 130 receives the window information from the Web server 20 (step S113) and transmits the information to the control unit 120. The control unit 120 transmits the window information to the window display unit 140 to display it (step S114).

The voice uttered by the user is input to the data input unit 110 of the client 10 (step S115) and transmitted to the control unit 120. The control unit

120 performs the data processing described in the above specific example. In this data processing, the unique ID is contained in the voice data.

The data communication unit 130 sequentially transmits the processed data to the voice processing server 30 (step S116). The processing of containing the unique ID in the voice data is the same as that in the above specific example.

In the voice processing server 30, the data communication unit 310 receives the sequentially transmitted data (step S311), and the control unit 320 determines that the received data is voice data, and transmits it to the voice processing executing unit 330.

In the voice processing server 30, the voice processing executing unit 330 comprises at least one of the following (not shown) required for voice processing (voice recognition, speaker collation, and the like) as in the above specific example: a recognition engine, recognition dictionary, synthesis engine, synthesis dictionary, and speaker collation engine, and sequentially performs voice processing (step S312). After the voice processing is complete, the voice processing executing unit 330 transmits the voice processing result to the data communication unit 310 through the control unit 320. The data communication unit 310 then transmits the result to the Web server 20 (step S313). At this point of time, the voice

processing server 30 transmits the unique ID to the Web

server 20.   The same transmission method as that in the

above specific example is performed.

In the Web server 20, the data communication

5   unit 210 receives the voice processing result and unique

ID transmitted from the voice processing server 30 (step

S214), and transmits them to the content management unit

220.

The content management unit 220 of the Web

10   server 20 prepares information reflecting the voice

processing result (second information: content

information corresponding to voice processing resultant

information and the voice processing result or content

information corresponding to the voice processing

15   result) in correspondence with the unique ID (step

S215).   The content management unit 220 of the Web

server 20 can discriminate the client 10 as the

transmission destination of the information reflecting

the voice processing result from the unique ID of the

20   client.

The Web server 20 then downloads, to the

client 10 which has issued the service request, the

resultant information (e.g., the voice recognition

identification result window on the window 1003 in

25   Fig. 15) and a content (e.g., a map window on the window

1004 in Fig. 15) which are generated for each unique ID

or only the content (e.g., the map window on the window

1004 in Fig. 15) (step S216), and the client 10 receives
the downloaded information (step S117). The information
is then displayed in a window on the client 10. The
same download method as that used in the above specific
example is used for the generated content information.
This specific example can be used for various kinds of
processing in, for example, the following cases:
·   a case wherein the client 10 interacts with the voice
processing server 30,
·   a case wherein the voice processing server 30
performs a search or the like by using the voice
processing result, or
·   a case wherein the Web server 20 performs a search or
the like by using the voice processing result.

            Note that the processing to be performed by
the client 10, Web server 20, and voice processing
server 30, which is exemplified by Fig. 13, i.e., the
functions of them, may be implemented by programs
executed on computers forming the client 10, Web server
20, and voice processing server 30.
[Third Specific Example]

            The third specific example of the present
invention will be described next. In this specific
example, the voice processing server 30 comprises a
processing unit which generates a session ID. Fig. 11
is a view showing the arrangement of the voice
processing server 30. Referring to Fig. 11, the voice

processing server 30 according to this specific example
is obtained by adding a session ID generating unit 340
to the voice processing server 30 shown in Fig. 8. The
client 10 according to this specific example has the

5   arrangement shown in Fig. 6, and the Web server 20 has
the arrangement shown in Fig. 10. The operation of the
specific example will be described below.

         Fig. 14 is a view for explaining the sequence
operation of this specific example. The specific

10   example will be described with reference to Figs. 6, 10,
11, and 14.

         The client 10 issues a request for a service
containing voice processing to the Web server 20 (step
S121).

15   On the Web server 20 side, the data
communication unit 210 receives the service request
signal (step S221) and transmits the signal to the
content management unit 220. The content management
unit 220 receives the service request signal and

20   generates a window for the requested service upon
checking the service (step S222), and transmits
(downloads) the window to the client 10 through the data
communication unit 210 (step S223).

         The client 10 receives the window information

25   from the Web server 20 (step S122), and transmits a
voice processing request signal to the voice processing
server 30 to transmit the voice information to the voice

processing server 30 (step S123).

In the voice processing server 30, the data communication unit 310 receives the voice processing request signal (step S321) and transmits the signal to the control unit 320. The control unit 320 transmits the voice processing request signal to the session ID generating unit 340.

The session ID generating unit 340 of the voice processing server 30 receives a session ID request signal and generates a session ID. The manner of generating a session ID is the same as that described in the above specific example.

The session ID generated by the session ID generating unit 340 of the voice processing server 30 is transmitted from the session ID generating unit 340 of the voice processing server 30 to the data communication unit 310 through the control unit 320.

The data communication unit 310 of the voice processing server 30 transmits the session ID to the client 10 (step S322).

The client 10 receives the session ID from the voice processing server 30 (step S124), and transmits the session ID to the control unit 120 through the data communication unit 130.

The session ID is then transmitted to the Web server 20 through the data communication unit 130 of the client 10 (step S125).

In the Web server 20, the data communication unit 210 receives the session ID (step S224), and transmits the session ID to the content management unit 220, thereby managing it.

After the client 10 notifies the Web server 20 of the session ID, the client 10 inputs the voice uttered by the user to the data input unit 110 (step S126), and transmits the voice to the control unit 120. The control unit 120 performs the same data processing as that in the above specific example. In this data processing, the session ID may be contained in the voice data.

The data communication unit 130 of the client 10 sequentially transmits the processed data to the voice processing server 30 (step S127).

In the voice processing server 30, the data communication unit 310 receives the data sequentially transmitted from the client 10 (step S323), and the control unit 320 determines that the received data is voice data, and transmits it to the voice processing executing unit 330.

The voice processing executing unit 330 comprises at least one of the following functions (not shown) required for voice processing (voice recognition, speaker collation, and the like) as in the above specific example: a recognition engine, recognition dictionary, synthesis engine, synthesis dictionary, and

speaker collation engine, and sequentially performs
voice processing (step S324). After the voice
processing is complete, the voice processing executing
unit 330 transmits the voice processing result to the
5   data communication unit 310 through the control unit
320. The data communication unit 310 then transmits the
result to the Web server 20 (step S325). The voice
processing result is handled in the same manner as in
the above specific example. At this point of time, the
10   voice processing server 30 also transmits the session ID
to the Web server 20. The session ID is transmitted in
the same manner as in the above specific example.

In the Web server 20, the data communication
unit 210 receives the voice processing result and the
15   session ID (step S225), and transmits them to the
content management unit 220. The voice processing
result has the same contents as those in the above
specific example. At this point of time, the voice
processing server 30 transmits the session ID to the Web
20   server 20 in the same manner as in the above specific
example. In the Web server 20, the data communication
unit 210 receives the voice processing result and
session ID and transmits them to the content management
unit 220.

25   The content management unit 220 of the Web
server 20 generates information reflecting the voice
processing result corresponding the session ID (content

information corresponding to the voice processing resultant information and the voice processing result or content information corresponding to the voice processing result) for each session ID (step S226).

5   The Web server 20 then downloads, to the client which has issued a service request, the resultant information (e.g., the voice recognition identification result window on the window 1003 in Fig. 15) and a content (e.g., a map window on the window 1004 in

10   Fig. 15) which are generated for each session ID or only the content (e.g., the map window on the window 1004 in Fig. 15) (step S226), and the client 10 receives the downloaded information from the Web server 20.

          The generated content information may be

15   downloaded by the following method. The client 10 executes the processing of setting the URL linked to the result acquisition button on the window downloaded to the client 10 to the URL contained in the session ID notified from the voice processing server 30 at the

20   start of voice processing. The Web server 20 then places the content information reflecting the voice processing result at the URL contained in the session ID. With this operation, when the user presses the result acquisition button (e.g., "display map" button on

25   the window 1003 in Fig. 15) on the client window, the content information reflecting the voice processing result is downloaded to the client 10.

As in the above specific example, this specific example can be used for various kinds of processing in, for example, the following cases:

· a case wherein the client 10 interacts with the voice processing server 30,

· a case wherein the voice processing server 30 performs a search or the like by using the voice processing result, or

· a case wherein the Web server 20 performs a search or the like by using the voice processing result.

Note that the processing to be performed by the client 10, Web server 20, and voice processing server 30, which is exemplified by Fig. 14, i.e., the functions of them, may be implemented by programs executed on computers forming the client 10, Web server 20, and voice processing server 30.

[Operation Window]

An example of an operation window on the client 10 as a specific example to which the present invention is applied will be described next. Fig. 15 is a view showing an example of transition to a window (page) displayed on the window display unit 140 of the client 10 in the first specific example of the present invention in which sequence operation has been described with reference to Fig. 12. Window display on the client 10 in the first specific example of the present invention will be described below with reference to

Figs. 15 and 12.

<Window 1001>

A window 1001 is a window (the top page of "map search") downloaded from the Web server 20, in which a CGI (e.g., http://····.jp/a.cgi) is linked to a "voice input" button 1011. When the user clicks the "voice input" button 1011 displayed on the window to issue a service request (corresponding to step S101 in Fig. 12), the Web server 20 activates a process (CGI program) called "a.cgi", and input information is transferred. The Web server 20 generates an HTML file on the basis of the processing result obtained by the CGI program and returns it as a response to the client 10.

<Window 1002>

A "voice input" window 1002 is displayed, and the message "please utter address on map for which you want to search like "Tokyo to minato ku mita" is displayed (which corresponds steps S102 to S104 in Fig. 12). An ID is embedded as a tag in the window. In this state of the window 1002, the user performs voice input (utterance). On the window, a page (http://····/b.ID.html) is linked to a "display result" button 1012. When the user clicks the "display result" button 1012 on the window, the recognition result obtained by

voice recognition performed by the voice processing
server 30 is displayed as indicated by a next window
1003. Note that the recognition result window displayed
on the window 1003 is the one downloaded from the Web
5   server 20 to the client 10.
<Window 1003>

The "recognition result" window 1003 is
displayed on the client 10 and the message "is result
"Tokyo to minato ku mita" ?" is displayed. A "display
10   map" button 1013 is display on the window.
<Window 1004>

When the user clicks the "display map" button
1013 on the window, content information is downloaded
from the Web server 20 (which corresponds to step S106
15   in Fig. 12), and a map window (page) 1004 is displayed.

Note that in this specific example, it
suffices to directly display the window 1004 as a result
of the window 1002 without displaying the window 1003 as
a recognition result window. That is, although the
20   voice processing server 30 generates the window 1003 as
a voice recognition result for each ID, the specific
example may be configured to directly display the window
1004 in which a voice recognition result is reflected by
clicking the "display result" button 1012 on the window
25   1002 (in this case, the window 1003 in Fig. 15 is

omitted).

Although Fig. 15 and Fig. 16 to be described
below each show an example of a window for a map
guidance system based on voice input, it is obvious that
5   the present invention is not limited to such a system,
and can be applied to arbitrary utterance management.

Fig. 16 is a view showing a modification of
Fig. 15. On a window 1002a shown in Fig. 16, the
"display result" button 1012 on the window 1002 in
10   Fig. 15 is not displayed. In the example shown in
Fig. 16, when voice input is performed on the window
1002a, the recognition result window 1003 is displayed
without clicking the "display result" button 1002a on
the window such as the window 1002 in Fig. 15. When the
15   user clicks the "display map" button 1013, the map of
the window 1004 is displayed. Alternatively, performing
voice input on the window 1002a directly displays a map
on the window 1004 without displaying the window 1003.

When a window is prepared for each ID (step
20   S206 in Fig. 12), the Web server 20 transmits the URL
information of the window to the client 10. The client
10 automatically access the received URL information.
As a consequence, the windows 1003 and 1004 shown in
Figs. 15 and 16 are displayed.

25   The procedure of processing in a case wherein

the user continuously utters voice on the client 10 in

the specific example of the present invention will be

described next. For example, when the user continuously

utter voice, a "voice re-input" button or the like may

5   be generated in advance on a window 1004 in Fig. 15 or

16. Linking the URL of a CGI of a Web server to the

"voice re-input" button on the window 1004 makes it

possible to generate a new ID when the user clicks the

"voice re-input" button on the window 1004, thereby

10   allowing the user to perform voice re-input operation.

Alternatively, a "to top page" button may be

generated in advance on the window 1004 in Fig. 15 or

16. When the user clicks the "to top page" button on

the window 1004, the page returns to the page of the

15   window 1001 in Fig. 15 or 16. This makes it possible to

perform the process of "voice input" again.

Obviously, it suffices to take countermeasures

such as using passwords or cryptography (public key

cryptography) as security measures such as security

20   protection for security IDs and unique IDs transferred

between the client 10, the Web server 20, and the voice

processing server 30.

The present invention has been described with

reference to the above specific examples. Obviously,

25   however, the present invention is not limited to only

the arrangements of the above specific examples, and incorporate various modifications and improvements made by those skilled in the art within the spirit and scope of the invention.

5   Industrial Applicability

The present invention is applicable to a service providing system which causes a client to display a window, issue a request by voice, and display the result. More specifically, for example, the present

10   invention can be suitably applied to

· a service of displaying a map in accordance with an address uttered in voice;

· a service of displaying a manual in accordance with a sentence to be searched for which is uttered in voice;

15   and

· a service of downloading a tune in accordance with the tile of the tune uttered in voice.

In addition, the present invention allows transmission/reception of data through a packet network,

20   and hence allows the user of a portable digital assistant (PDA), PC, on-vehicle terminal, home terminal, and the like as clients as well as a cell phone terminal.